DOCUMENT RESUME

ED 229 410 TM 830 308

AUTHOR Diamond, Esther E.
TITLE Development of the Joint Committee Standards for Evaluations of Educational Programs, Projects, and Materials.
PUB DATE Apr 83
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Committees; Educational Quality; *Evaluation Criteria; *Evaluation Methods; *Evaluation Needs; History; Program Development; *Program Evaluation; Research Problems; *Standards
IDENTIFIERS *Joint Committee on Standards for Educ Evaluation; Standards for Evaluation Educ Prog Proj Materials

ABSTRACT

The growing demand for program evaluation in the 1970s gave rise to a need for a comprehensive, carefully developed, objective set of guiding principles for the evaluation process, from initial planning to final report. The Joint Committee on Standards for Education Evaluation was established to meet this need. This broad-based group, representing 12 organizations with widely differing constituencies, spent approximately 6 years developing the "Standards for Evaluation of Educational Programs, Projects, and Materials," published by McGraw-Hill in 1981. The intended audience includes all who commission, conduct, or employ the results of evaluations to improve education. In developing the standards, the Joint Committee strove for a set of general rules for dealing with a variety of specific evaluation concerns, and a conceptual framework appropriate to a broad spectrum of such concerns. At the same time, the Committee considered it essential that the standards promote innovation rather than stifle it, that research on their applicability and the effectiveness of their use be encouraged, that training needs be addressed, and that revisions be planned as needed. An overview of the development process and of the 30 standards, including their rationale and organization and current issues regarding their use is provided. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED229410

Development of the Joint Committee Standards for Evaluations of Educational Programs, Projects, and Materials[1]

Esther E. Diamond

Educational and Psychological Consultant

Evanston, Illinois

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E E Diamond

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running head: Joint Committee Standards

[1]Paper presented at symposium, Professional Standards for Evaluation: An Analysis of Similarities and Differences, at the annual meeting of the American Educational Research Association, Montreal, Canada, April 1983.

No portion of this paper may be distributed or reproduced in any way without prior written permission of the author.

TM 830 308

Printed in U.S.A.

Development of the Joint Committee Standards for Evaluations of Educational Programs, Projects, and Materials

Esther E. Diamond
Educational and Psychological Consultant
Evanston, Illinois

Abstract

The growing demand for program evaluation in the 1970s gave rise to a number of critical questions: How effective were the evaluations? What were the evaluation criteria? Was evaluation helping to improve program quality? Clearly, there was need for a comprehensive, carefully developed, objective set of guiding principles for the evaluation process, from initial planning to final report. The Joint Committee on Standards for Educational Evaluation was established to meet this need. This broad-based group, representing twelve organizations with widely differing constitutencies, spent approximately six years developing the Standards for Evaluations of Educational Programs, Projects, and Materials, published by McGraw-Hill in 1981. The intended audience includes all who commission, conduct, or employ the results of evaluations to improve education. In developing the Standards, the Joint Committee strove for a set of general rules for dealing with a variety of specific evaluation concerns, and a conceptual framework appropriate to a broad spectrum of such concerns. At the same time, the Committee considered it essential that the Standards promote innovation rather than stifle it, that research on their applicability and the effectiveness of their use be encouraged, that training needs be addressed, and that revisions be planned as needed. An overview of the development process and of the thirty standards, including their rationale and organization and current issues regarding their use, is provided.

Development of the Joint Committee Standards for Evaluations of Educational Programs, Projects, and Materials

Esther E. Diamond
Educational and Pyschological Consultant
Evanston, Illinois

The story of the development of the Joint Committee Standards for Evaluations of Educational Programs, Projects, and Materials really begins with the evolution of evaluation over the past fifty to sixty years. In the early years, evaluation was associated mainly with assessment of student performance. But in the last fifteen to twenty years, with the advent of federally funded programs designed to make the most of U.S. talents and improve the quality of education at all levels, Congress began requiring evaluations of projects, programs, and materials funded by the federal government--not only in education, but in health, social services, and other areas as well. Inevitably, questions were raised about the evaluations themselves, and soon the need for standards for program evaluation became apparent.

Controversial Issues in Evaluation

Evaluations of the '60s and '70s were fraught with problems and beset by controversy. For example, to what extent should evaluators be involved with the goals of a project or program? Scriven (1972) argued in favor of "goal-free" evaluation: The evaluator should not be involved with program goals, but should focus on the product itself--the effects of the program, including the side-, or unanticipated, effects, vs. demonstrated needs.

Another common problem was an evaluation dilemma described by Rein and White (1978)--that of the policy-research alliance, with science seeking to establish value-neutral patterns of experience acceptable to all "whether they find them palatable or unpalatable" (p.40). Politics, on the other hand, seeks one purpose or course of action acceptable to individuals espousing different purposes, courses of action, and values. It is difficult to maintain credibility in these two almost diametrically opposed areas.

A prime example of problems in evaluation is the ten-year $30 million-plus Follow-Through evaluation effort (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977). In this evaluation and the controversies that followed it, the program was viewed somewhat in the manner of the blind men describing an elephant. As House, Glass, McLean, and Walker (1978) pointed out, the U.S. Office of Education (USOE) viewed the program as an experiment to determine which of a number of models for compensatory education for disadvantaged children worked best. Sponsors of the individual models, on the other hand, viewed the effort as program development, and parents saw it as a service for their children. These diverse perceptions led to conflicting expectations of results. The Stanford Research Institute, responsible for data collection, and the USOE had promised to develop measurement instruments specifically designed to measure certain kinds of outcomes considered important by sponsors and parents, but then USOE decided to cut back drastically on data collection, and only a few standardized instruments were used. The evaluation, which found no significant between-model program effects, made sponsors and parents--especially black parents, who were large stakeholders in the program--unhappy; the

feeling was widespread that the evaluation was not responsive to their concerns.

House et al. raised the question whether "What works best?" was the appropriate question to pursue, or whether a quasi-experimental design would have worked better. They pointed out various confounding effects: Models varied widely, from highly structured to very open systems, and the variations within each of the thirty models was also considerable; intersite variation among models was much larger than between-model differences. There were other problems: Outcome measures assessed very few of the models' goals and favored models that concentrated on teaching technical skills; the statistical procedure (analysis of covariance) and the unit of analysis (the individual pupil) were, House et al. felt, questionable; unique features of local settings complicated the problem further; and models that worked well in one place did not work in another. House et al. made a series of recommendations, including one that fairness should be a major criterion for judging an evaluation: This involves fulfilling promises of what will be done, and treating all parties to the evaluation equitably.

Whether House et al.'s critique is justified or not is not the question here. The original evaluation and the critique are cited here to illustrate the problems encountered in evaluation and to point up the need for evaluation criteria.

A more recent controversy was triggered by the Holtzman report (Boruch & Cordray, 1980), commissioned by the Department of Education as a result of legislation requiring a comprehensive review of federal evaluation practices and procedures. Among the report's recommendations

was use of high-quality evaluation designs, especially randomized experiments, which Bunda (1981) criticized because many evaluators regard experimental methods as providing "narrow answers to broad questions" (p. 14) and prefer the case study method. No single methodology, Bunda contended, should be singled out as high quality "which by implication leaves all others at a lower level. Evaluators should be charged with matching the methodology that is compatible with the question that they are addressing and the needs of the client" (p. 14). Bunda's second point of disagreement was with the emphasis on federal level with respect to interaction with clients and the focus of evaluation questions, which could detract from the feeling of local responsibility.

An article by Eichelberger (1979), which appeared when the Joint Committee had already produced a draft form of its Standards, identifies four issues that sum up the kinds of concerns the Joint Committee dealt with:

1. Most evaluation situations involve contending stakeholders with diverse values and interests in the evaluation. Results of the evaluation must be credible to the various stakeholder groups, who usually have different idealogical, theoretical, and practical perspectives.
2. All data involve some degree of subjectivity in their collection and interpretation, and each group can be expected to use whatever information is available to support its own position, regardless of the quality of the data.
3. Meaningful interpretation of evaluation results requires knowledge of the assumptions on which they are based. Basic assumptions of an experimental design are often not met in non-laboratory settings, where it is often difficult to shield variables against contamination,

to match traditional treatment groups on all but the mode of treatment, and so on.

4. Programs are dynamic in nature, and their evaluations must be flexible enough to change in order to meet new decision needs.

The Joint Committee Underway

The Joint Committee's work began when an early draft of the 1974 revision of the APA/AERA/NCME Standards for Educational and Psychological Tests (American Psychological Association, 1974) included a section on program evaluation and the use of tests for that purpose. The Committee on Test Standards expressed several reservations--time needed to develop adequate standards for evaluation; problems associated with evaluation that were eyond the scope of the committee's charge; the complexity of the field of evaluation and the need to develop standards directed toward a variety of specific settings; and the concern that the audiences for the two sets of standards--those involving tests and those involving program evaluation--were sufficiently different to require two separate sets of standards.

A memorandum from George Madaus, a National Council for Measurement in Education (NCME) representative on the Committee on Test Standards, urged preparation of a companion volume of program evaluation standards. After due deliberation, the three organizations agreed and appointed the initial Joint Committee on Standards for Educational Evaluation, charging them to prepare a proposal for such a project. At their first meeting, in May 1975, the committee decided that the focus of their effort should be broadened beyond the specific issue of the use of tests in evaluation; that their initial effort should concentrate on evaluation in education; and that the Joint Committee should be expanded to represent the various groups involved in evaluating education. The audience was to be those who commission, conduct, or

employ the results of evaluations to improve education--including professional educators, legislators, school board members, parents, and others.

The broadened Joint Committee was to have eighteen members, including nine representing groups concerned with the improvement of evaluation methodology and nine representing groups concerned with the improved application of evaluation methodology./A draft proposal was appproved in December 1975, Daniel Stufflebeam was elected committee chair, and he undertook the search for financial support for the project. Actually, there were twelve sponsoring organizations.

In April 1976 the funded project really got underway; grants had been obtained from the Lilly Endowment, the National Institute of Education, the National Science Foundation, and the Weyerhaeuser Company Foundation. The sponsoring organizations also contributed funds. The project staff was housed at the Western Michigan University Evaluation Center. Four categories of standards were identified, and a list of topics for each was developed, as well as a format for writing a standard. A nationwide panel of evaluation experts was selected, and each member was commissioned to write several standards. The Joint Committee and the project staff then critiqued and rewrote the standards. The process from then on was one of continuous development, review, refinement, and further review--through four successive drafts. A series of field tests tested the applicability of the standards. Hearings were held across the country. A final draft was developed in winter of 1980, and published by McGraw-Hill in 1981 as Standards for Evaluations of of Educational Programs, Projects, and Materials. The book is now in its second printing.

Throughout, the process involved the reconciling of a variety of viewpoints through lengthy dialogue, intensive small-task-group work, repeated

trips back to the drawing board, and--all-in-all--an exemplary model of group process at work. While the diversity of involvement undoubtedly added to the time and cost of the process, it also undoubtedly enhanced the quality and credibility of the effort; the Standards take into consideration the diversity in scope and nature of evaluations and have avoided locking evaluators into a rigid set of do's and don'ts.

The Joint Committee viewed evaluation as the systematic assessment of the worth or merit of the object of the evaluation, and it recognized that there are a number of ways of assigning value meaning to findings. Its members agreed that no adequate standards for educational evaluation existed, and that a set of professional standards could play a vital role in upgrading educational evaluations. At the same time, the Committee agreed that "no set of standards should be enshrined as complete and adequate for all time" (p.6) and that the Standards should be regarded as a set of guiding principles rather than mechanical rules. It also strove for a common language of evaluation and a conceptual framework appropriate to a broad spectrum of evaluation concerns.

Organization of the 30 Standards

The thirty standards are organized into four categories that represent four principal concerns about any evaluation--its utility, feasibility, propriety, and accuracy.

Utility standards are standards for guiding evaluations so that they will be informative, timely, and influential; they are concerned with whether an evaluation serves the practical information needs of a given audience.

Feasibility standards call for evaluations to be realistic, prudent, diplomatic, and frugal. They recognize that educational evaluations are

usually conducted in a natural rather than a laboratory setting and that they must not consume more materials and personnel than necessary to achieve their purposes.

Propriety standards reflect the fact that educational evaluations affect many human beings in many ways and that the rights of persons so affected must be protected. These standards warn against unlawful, unethical, or inept actions by those who produce evaluation results.

Accuracy standards are designed to determine whether an evaluation has produced sound information and an accurate, comprehensive assessment of the object being evaluated so that its merit or worth can be appropriately determined.

Within these four groups, the standards are presented in a common format:

- a brief description
- a definition in the form of a "should" statement
- an overview, or conceptual/introductory statement that gives definitions of key terms in the standard and a general rationale for its use
- a list of guidelines or procedural suggestions intended to help evaluators and their audiences to meet the requirements of the standard
- a list of not easily recognized difficulties or pitfalls believed to be associated with the particular standard
- caveats about mistakes based on overzealousness in application of the standard, and a discussion of the trade-offs that may be necessary in applying the standard in different situations

- an illustrative case showing how the standard might be applied, a situation in which the standard is not met, and a discussion of corrective actions that would result in the standard being met

Taken as a whole, the thirty standards are a proposed working philosophy of evaluation.

Some Considerations Regarding Applicability

In the introduction to the Standards, the Joint Committee discusses some of the limitations that might be encountered in applying them, and important points to keep in mind. For example, the importance of a standard at any one of a series of suggested checkpoints depends on the context of the evaluation. The relative importance of individual standards will differ from one context to another, and sometimes one or more standards will conflict with one or more others. The Joint Committee recommends that all standards be considered equally important until they can be ranked in the context of a particular situation. Conflicting standards should also be viewed within such context and necessary tradeoffs worked out as judiciously as possible, whether it means delaying a needed report in order to produce valid and reliable information or reducing dissemination plans somewhat in order to stay within cost limits.

Applicability of the standards and their limitations were discussed in a special feature of Evaluation News (1981). Sechrest (1981) found the standards demanding, and argued that few evaluators will be able to live up to them. He also pointed out what the Joint Committee was already aware of: Some of the standards might actually be incompatible with each other. He also expressed disappointment that there was no reference in the Standards to appropriate experimental or quasi-experimental designs

for evaluation-- a reflection of the Joint Committee's decision not to espouse any one type of methodology or analysis over another.

In the same feature, Stake (1981) raised several issues, mainly building a case against any standards at all. The profession, he argued, should attempt to control the individual evaluator only when his or her actions place individuals in jeopardy--that is, introduce adverse impact. The very existence of standards, he maintained, brings injustice and constraints on creativity, stifling development of new techniques and viewpoints. Explicit standards, moreover, Stake argued, "legitimate everything they do not explicitly prohibit" (p. 151), and the best standards are those that are internal within individuals and implied within their work. The Joint Committee, throughout the process of developing the standards, has painstakingly endeavored to avoid or reject any recommendation that might even remotely result in stifling of creativity or in acceptance of only a single, narrow viewpoint. Only the test of time and use will tell whether their efforts have been successful and whether Stake's concerns are without foundation. Such concerns are bound to arise with any set of standards or rules of the game applicable to many fields in addition to that of evaluation. In the final analysis, one must judge whether the advantages outweigh the possible dangers, and then one must somehow warn against the possible pitfalls.

Next Steps

Research on the applicability and effectiveness of use of the Standards on an ongoing basis is a major priority of the Joint Committee. Training needs in the use of the Standards are being addressed on a nation-wide basis. Revisions will be planned as needed in line with future developments in evaluation and with findings from research studies. At its last meeting,

in October 1982, the Committee voted to explore the feasibility of broadening its base beyond educational evaluation. As Rossi (1982) commented:

> Indeed, one may ask why the Joint Committee addressed itself only to educational evaluation, since the standards involved are not particular to educational evaluations. Very little would have been lost had the Joint Committee addressed itself to general standards applicable to all substantive areas. (p. 3)

A subcommittee of the Joint Committee is now at work on the task of making the Standards applicable to all substantive areas.

References

Boruch, R.F., & Cordray, D.S. (Eds.). An appraisal of educational program evaluations: Federal, state, and local agencies. Report to the Congress. Washington, D.C : Office of the Assistant Secretary for Management. U.S. Department of Education, 1980.

Bunda, M.A. Some comments on the recommendations in the Holtzman project. Educational Researcher, 1981, 10 (4), 14.

Eichelberger, R.T. Evaluating educational programs: Some basic considerations and criteria. The CEDR Quarterly, 1979, 12, 16-19.

House, E.R., Glass, G.V., McLean, L.D., & Walker, D.F. No simple answer: Critique of the Follow Through evaluation. Harvard Educational Review, 1978, 48, 128-160.

Joint Committee on Standards for Educational Evaluation. Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill, 1981.

Rein, M., & White, S. Can policy research help policy? In T.D. Cook (Ed.), Evaluation Studies Review Annual, Vol. 3. Beverly Hills, Calif: Sage Publications, 1978.

Rossi, P.H. Editor's notes. In P. Rossi (Ed.). New Directions for program Evaluation: Standards for Evaluation Practice, no. 15. San Francisco: Jossey-Bass, 1982.

Scriven, M. Prose and cons about goal-free evaluation. In W.J. Popham (Ed.), Evaluation in education. Berkeley, Calif: McCutchan Publishing, 1974.

Sechrest, L. The Standards: A general review. Evaluation News, 1981, 2 145-148.

Special feature: An examination of the Standards for evaluations of educational programs, projects, and materials. Evaluation News, 1981, 2, 140-170.

Stake, R.E. Setting standards for educational evaluation. Evaluation News, 1981, 2, 148-152.

Stebbins, L.B., St. Pierre, R.G., Proper, E.C., Anderson, R.B., & Cerva, T.R. Education as experimentation: A planned variation model, Volume IV-A, An Evaluation of Follow Through. Report to the U.S. Office of Education pursuant to Contract No. 300-75-0134. Cambridge, Mass.: Abt Associates, 1977.